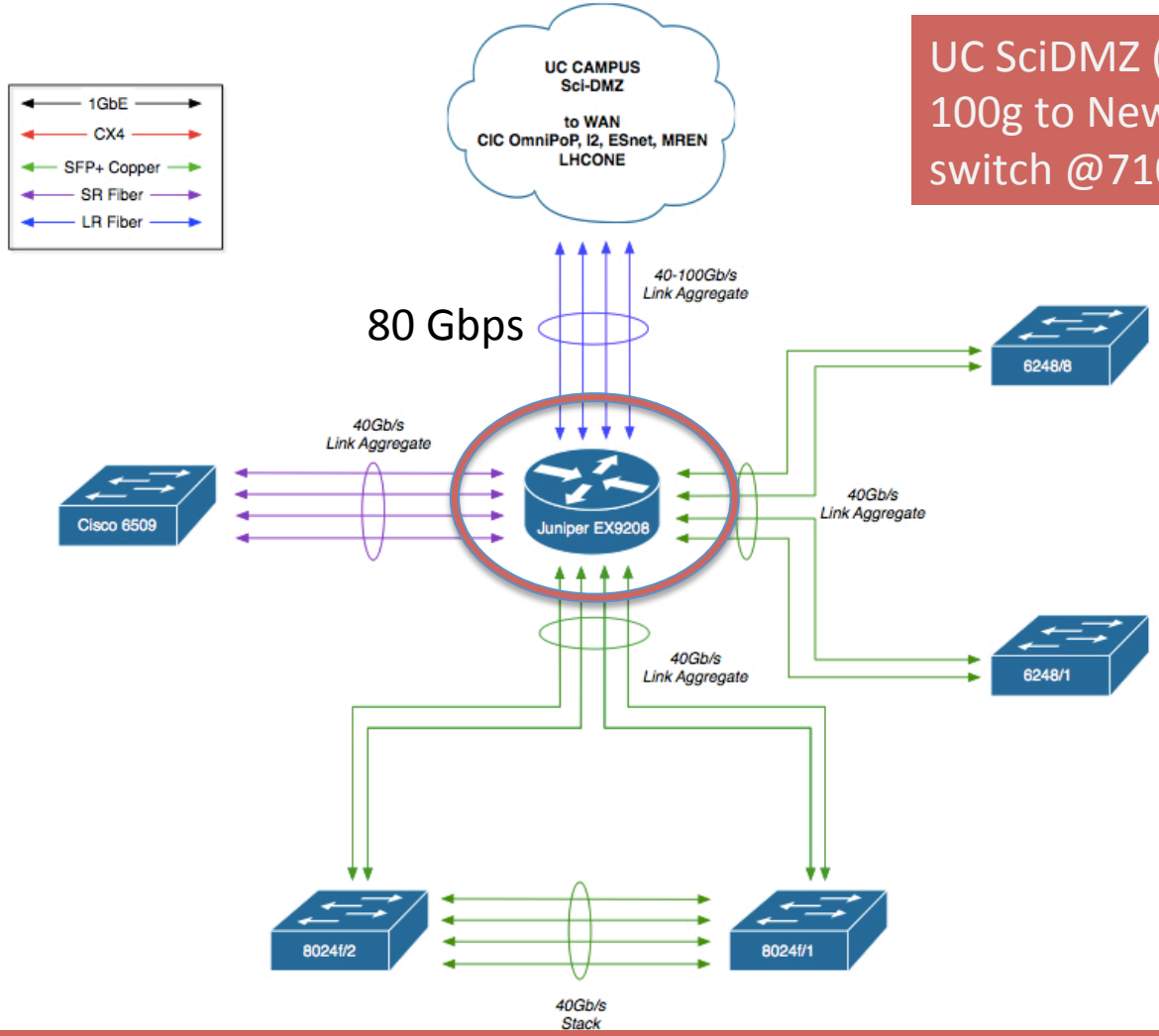
# Request for Beyond Target Funding at UChicago to improve local and WAN IO performance

Rob Gardner, Lincoln Bryant
6/19/2014

# Background

- Previously UChicago was awarded a RBT of $125k for networking upgrades at the Tier2 center.
- This was used to purchase a 100 Gbps capable Juniper EX9208 router, optics, and cables.
- Now have 80 Gbps capacity uplink from the Tier2 to the WAN via a SciDMZ
  - Provided by 2013 NSF CC-NIE funds
  - Good WAN connectivity to ESNet, I2, BNL, between the MWT2 sites, AGLT2, etc., via CIC OmniPoP at Starlight
- Reconfigured LAN with Juniper as central aggregation switch (40 Gbps LAG trunks to compute and storage switches)

# MWT2 UChicago Network Upgrade 2013-14



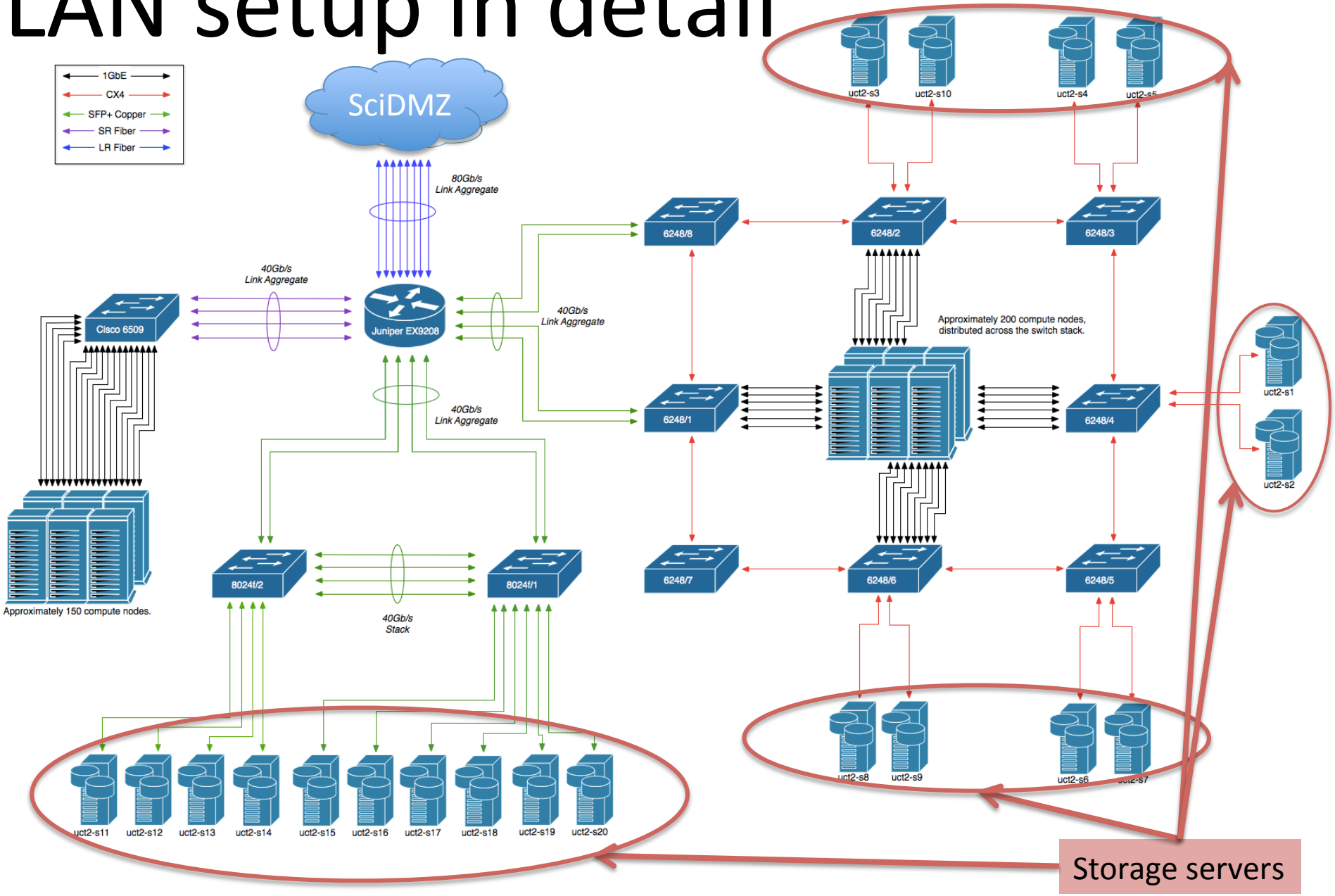

UC SciDMZ (NSF CCNIE funds): 100g to New CIC OmniPoP switch @710 NLSD

RBT program funds purchased Juniper EX9208 plus optics, cables: $125k

Initial connectivity @ 4x10g
WAN now 80g

**LAN reconfigured**: the EX9208 is the aggregation hub to 4 x 40Gbps trunked stacks (Cisco 6509, PC8024F stack, two 6248 stacks - one not shown)

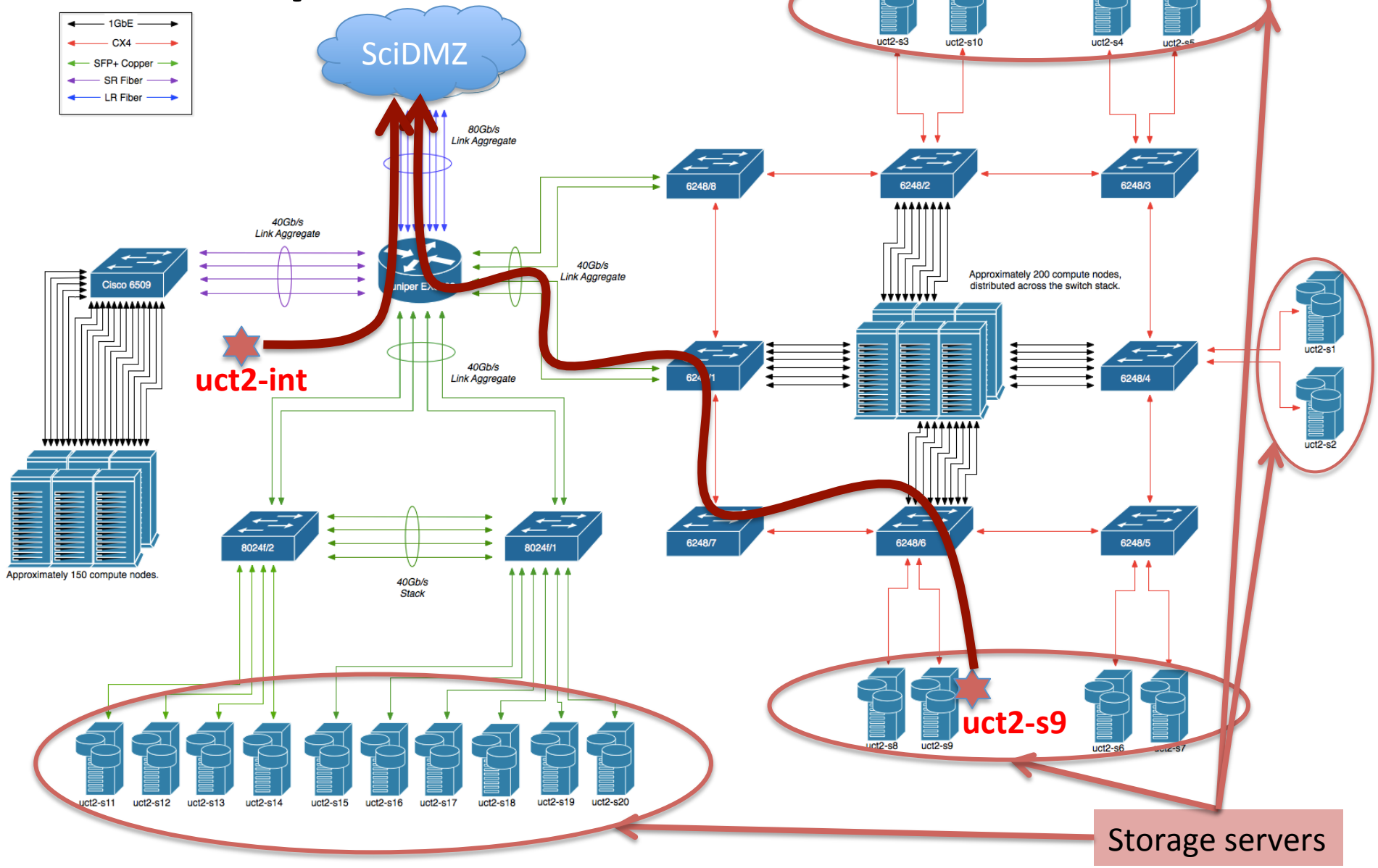
# LAN setup in detail



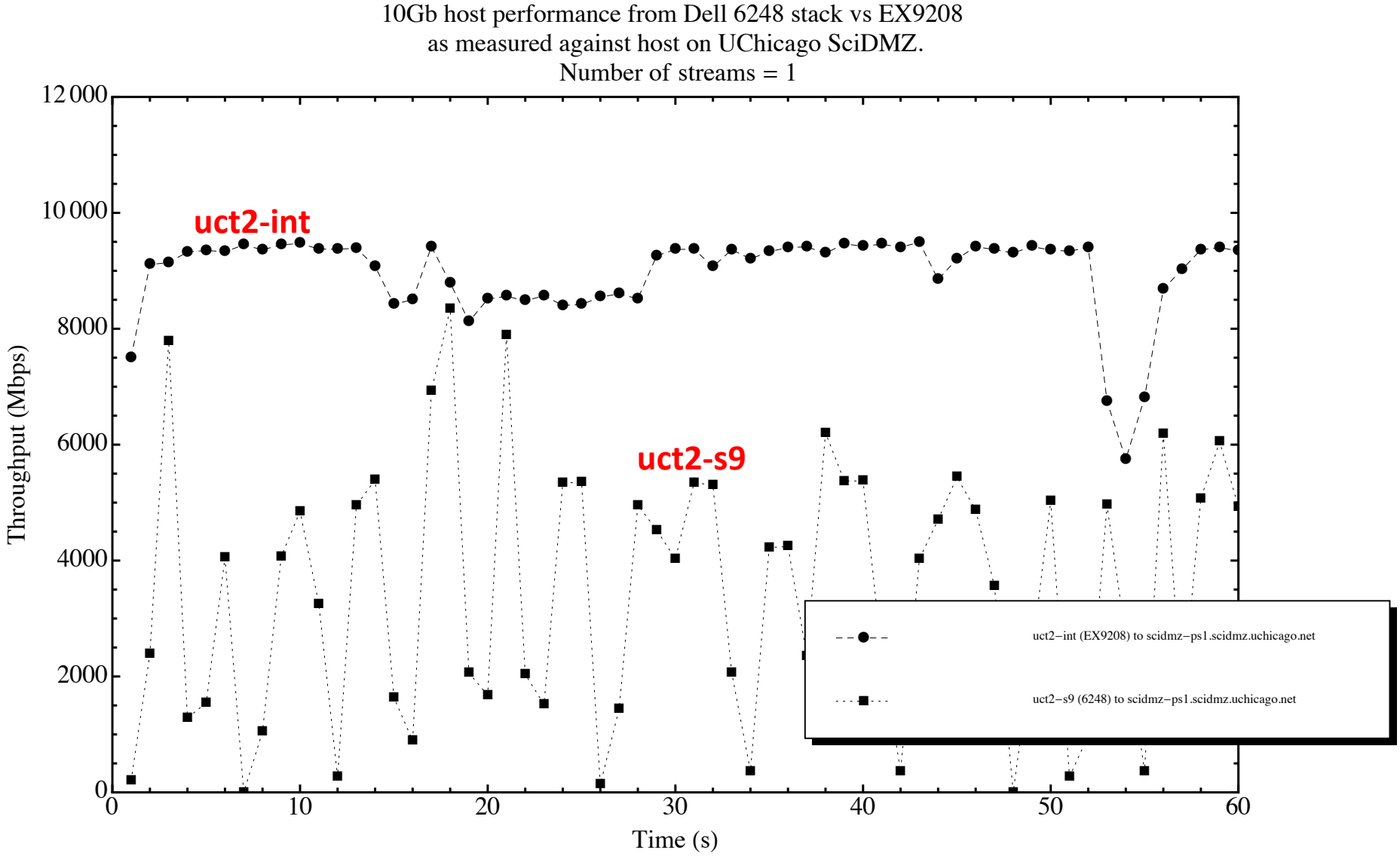
# Throughput Measurements

- **Test 1: Tier2 to SciDMZ host (single stream)**
  - Single stream TCP performance between 10g host directly attached to Juniper (uct2-int) and 10g host at campus edge (still within the SciDMZ)
  - Same, but for a 10g storage server (uct2-s9) attached to the PC6248 switch
- Paths and plot follow on next two slides
- Result: The node on the 6248 acts erratically. Since this is just a single stream, we expect a general dip in performance since we can only send on a single interface at a time and these interfaces are otherwise loaded with normal traffic. That is to say, we can only take advantage of 1 of our bonded interfaces instead of all 8 on the EX9208 -> SciDMZ uplink.

# Test 1 paths


1GbE
CX4
SFP+ Copper
SR Fiber
LR Fiber
SciDMZ
80Gb/s
Link Aggregate
40Gb/s
Link Aggregate
40Gb/s
Link Aggregate
40Gb/s
Link Aggregate
40Gb/s
Stack
Cisco 6509
Approximately 150 compute nodes.
8024f/2
8024f/1
6248/8
6248/2
6248/3
6248/7
6248/6
6248/5
6248/4
Approximately 200 compute nodes,
distributed across the switch stack.
uct2-int
uct2-s3
uct2-s10
uct2-s4
uct2-s5
uct2-s1
uct2-s2
uct2-s8
uct2-s9
uct2-s6
uct2-s7
uct2-s9
uct2-s11
uct2-s12
uct2-s13
uct2-s14
uct2-s15
uct2-s16
uct2-s17
uct2-s18
uct2-s19
uct2-s20
Storage servers

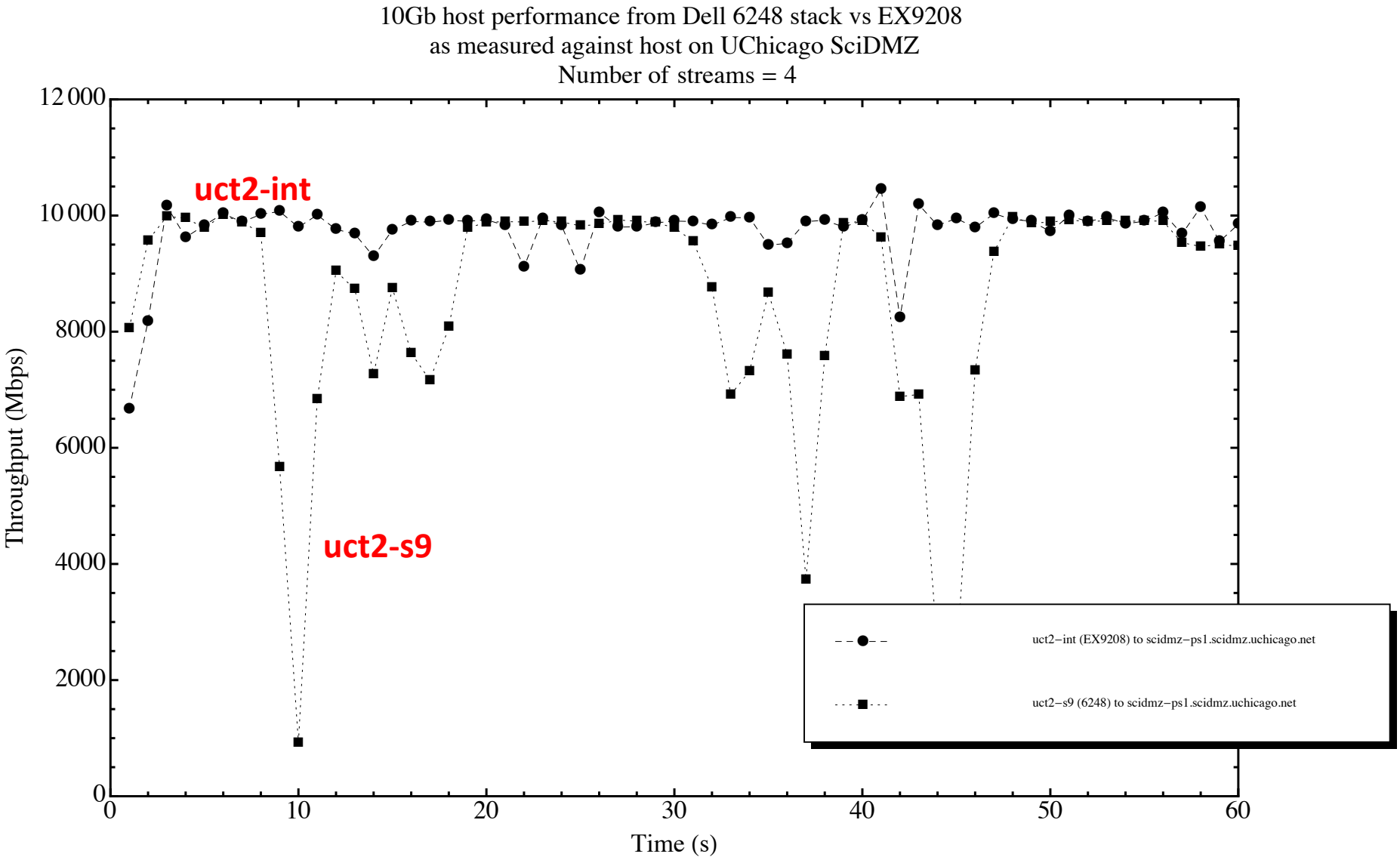
# Test 1 Result (single stream)



# Throughput Measurements (2)

- **Test 2: Tier2 to SciDMZ host (4 streams)**
  - Same as Test 1, but with four TCP streams
- Plot follows (paths same)
- Result:
  - Since we have 4x10Gbps interfaces bonded on the EX9208 <-> 6248 connection, the test should be load balanced between them for an aggregate ~40Gbps of available bandwidth. As you can see, the node on the 6248 performs better but still rather erratically. The node sitting directly on the EX9208 is operating very close to line speed.

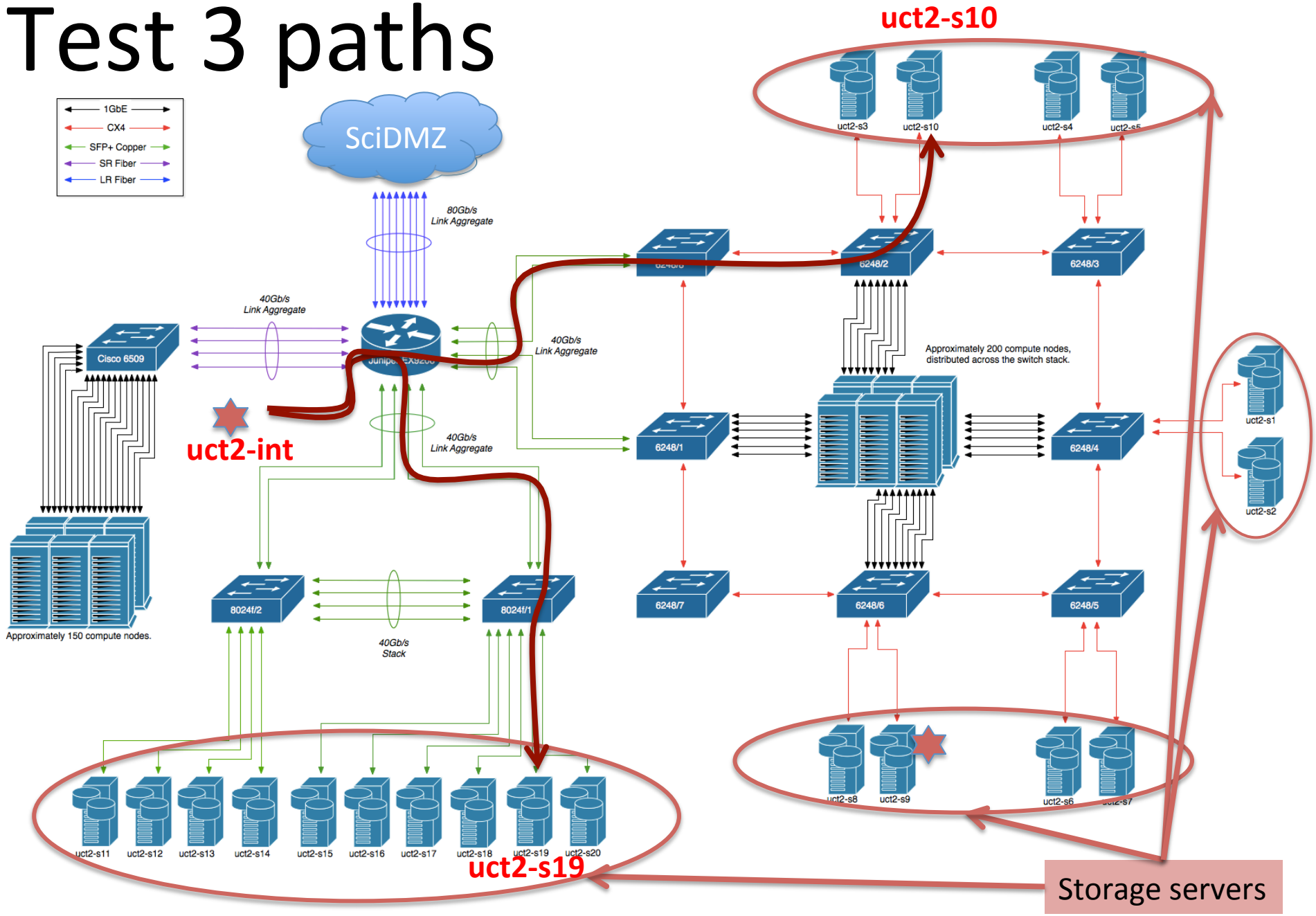
# Test 2 Result (4 streams)



# Throughput Measurements (3)

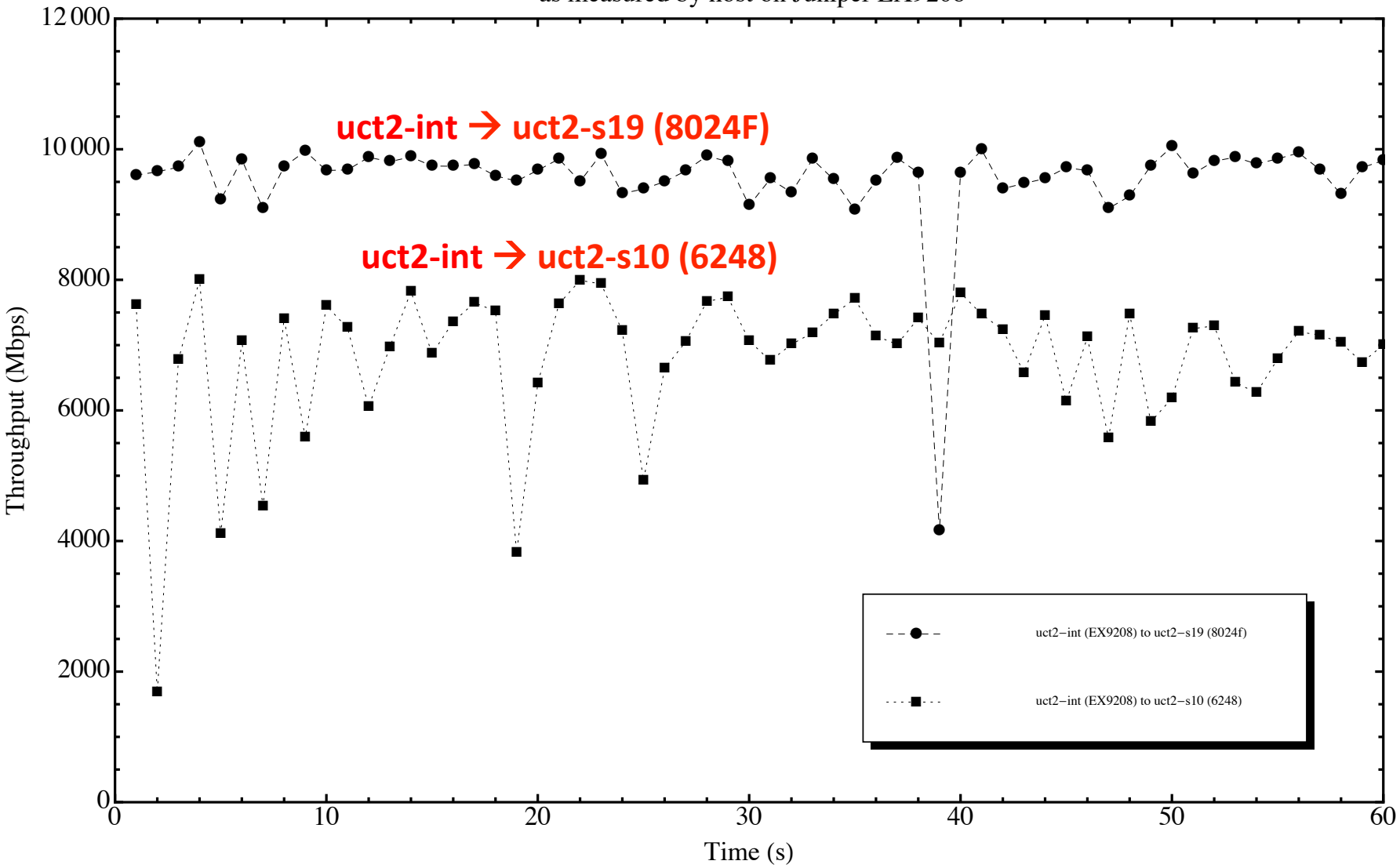
- **Test 3 LAN PC6248 vs. PC8024 via EX9208**
  - From the 10g host directly attached to EX9208 (uct2-int), probe 10g servers on the PC6248 (uct2-s10) & PC8024F (uct2-s19)
- Paths and plot follow

# Test 3 paths


1GbE
CX4
SFP+ Copper
SR Fiber
LR Fiber
SciDMZ
80Gb/s
Link Aggregate
40Gb/s
Link Aggregate
40Gb/s
Link Aggregate
40Gb/s
Link Aggregate
40Gb/s
Stack
Cisco 6509
Approximately 150 compute nodes.
8024f/2
8024f/1
6248/1
6248/2
6248/3
6248/4
6248/5
6248/6
6248/7
Approximately 200 compute nodes,
distributed across the switch stack.
uct2-int
uct2-s10
uct2-s3
uct2-s10
uct2-s4
uct2-s5
uct2-s1
uct2-s2
uct2-s8
uct2-s9
uct2-s6
uct2-s7
uct2-s11
uct2-s12
uct2-s13
uct2-s14
uct2-s15
uct2-s16
uct2-s17
uct2-s18
uct2-s19
uct2-s20
uct2-s19
Storage servers

# Test 3 Results

10Gb host performance from Dell 6248 stack vs Dell 8024 stack
as measured by host on Juniper EX9208



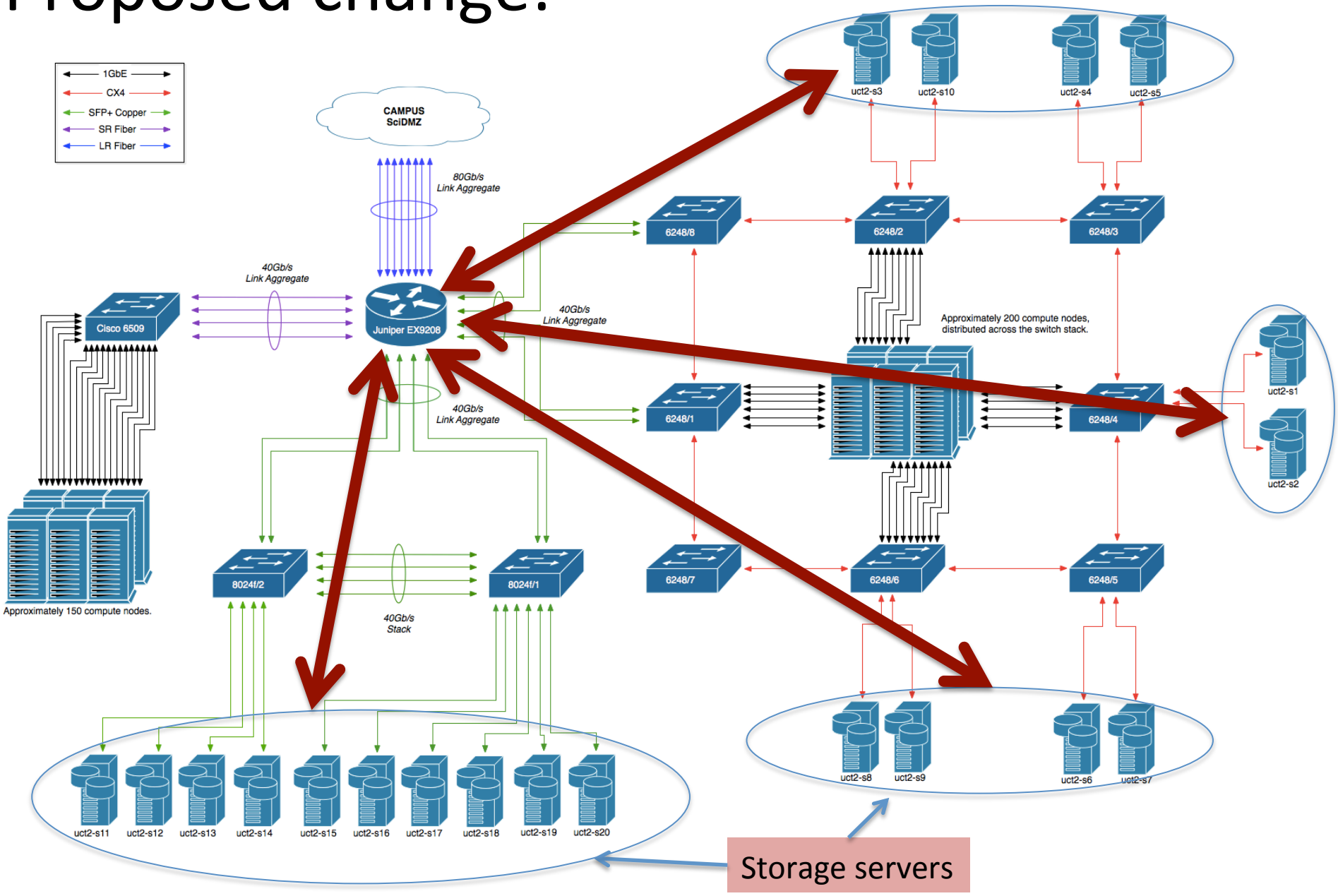
# Throughput Measurements (3), cont.

- Result:
  - As you can see, compared to the 8024 the host sitting on the 6248 is far more erratic and on average slower by 2Gbps
  - Thus we want to get our storage off the 6248 as both WAN and LAN are degraded

# Suggested Change (Centralizing Storage)

- Take advantage of the high backplane capacity of the EX9208 (4.6 Tbps)
- Move our 20 storage servers directly to the Juniper, each at 20 Gbps (2x10 Gbps)
- Thus storage servers are attached to our best switching equipment
- Removes a layer of switching for our newest storage servers (8024F)
- Also brings storage closer to the edge for WAN accesses, so better FAX and WebDAV performance
- Also allow wider trunks to the switch stacks serving the compute nodes

# Proposed change:



# Networking RBT Request

- Purchase 2 additional 32 port 10 Gbps line cards for the existing Juniper
- 2x EX9200-32XS
  - **Description:** JUNIPER NETWORKS 32PORT 1GBE/ 10GBE SFP/SFP+ LINE CARD REQUIRES SFP/SFP+ OPTICS
  - **Unit Price:** $24,000.00 **Subtotal:** $48,000.00
- UC would cover the cost of SFP+ optics, cables, and server NICs

**Quote for Security Solutions**

Nexum, Inc.
190 South LaSalle Street
Suite 1450
Chicago, IL 60603
+1-312-726-6900
www.nexuminc.com

| | |
|---|---|
| **Quote Number** | 2014-20867 |
| **Issue Date** | 1/21/2014 |
| **Expiration Date** | 2/28/2014 |
| **Customer ID** | UOC |
| **Payment Terms** | Net 30 |

**Quote To**
The University of Chicago
Ryan Harden
6045 S. Kenwood
Chicago, IL 60637
USA

**Deliver To**
University of Chicago
Ryan Harden
6045 S Kenwood
Chicago, IL 60637

**Sales Contact: Dirk Nowka**

**p.** (312) 506-6132 **m.** +1 3123435658
**f.** (312) 726-4451 **e.** dirk@nexuminc.com

**Inside Sales Contact: Alex Bangs**

**p.** (312) 506-6152 **e.** alex@nexuminc.com

**Juniper Networks**
***Pricing Reflects Pending Approval From Juniper***

| Qty | Part Number | Description | Unit Price | Ext. Price |
|---|---|---|---|---|
| 2 | EX9200-32XS | JUNIPER NETWORKS 32PORT 1GBE/10GBE SFP/SFP+ LINE CARD REQUIRES SFP/SFP+ OPTICS | $24,000.00 | $48,000.00 |

**Subtotal:** $48,000.00

| | |
|---|---|
| **Subtotal:** | **$48,000.00** |
| **Total:** | **$48,000.00** |